

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平6-332778

(43) 公開日 平成6年(1994)12月2日

(51) IntCl.⁶

G 0 6 F 12/00

識別記号

5 3 1 R 8944-5B

庁内整理番号

F I

技術表示箇所

審査請求 有 請求項の数 6 O L (全 17 頁)

(21) 出願番号 特願平6-86114

(22) 出願日 平成6年(1994)4月25日

(31) 優先権主張番号 0 6 6 3 6 0

(32) 優先日 1993年5月21日

(33) 優先権主張国 米国 (U S)

(71) 出願人 390009531

インターナショナル・ビジネス・マシー
ズ・コーポレーションINTERNATIONAL BUSIN
ESS MACHINES CORPO
RATIONアメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72) 発明者 チャンドラセカラン・モーハン

アメリカ合衆国95120 カリフォルニア州
サンノゼ ポストウッド・ドライブ727

(74) 代理人 弁理士 合田 潔 (外2名)

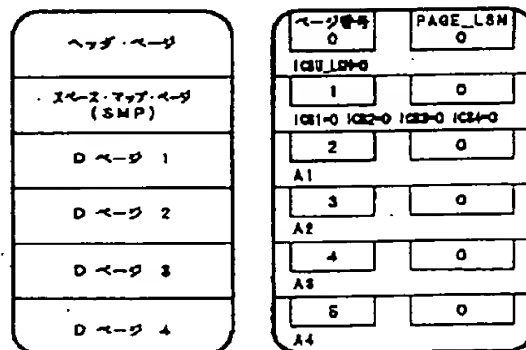
最終頁に続く

(54) 【発明の名称】 トランザクション管理方法

(57) 【要約】

【目的】 ログ・ベースの段階式コミット・トランザクション管理システム (TMS) において、修正可能なページのファイルをアーカイブ保存するための方法を提供すること。

【構成】 これは、1対の大域ログ・シーケンス番号の管理によって達成される。第1の番号 (ICBU_LSN) と各データ・ページLSNをページが修正されるときに比較することにより、共通状況ページを、変化した状況を正確に反映するように更新することができる。同じページへのその後の修正には、状況ページの訂正を必要としない。状況ページ標識は、バックアップ手順の一環として、増分式複写のためにページ・コピー・セットを確認するためにリセットされる。第2の番号 (ICRF_LSN) は、ファイルの復元に使用される。この場合、ICRF_LSNが、最近のコピーが作成されて以降の、ログ内の再実行のための点を定義する。



ファイル "A"

ファイル "A" 内のページ

ファイル名	日付	時間	完全または増分式	ファイル・コピーのテープ・アドレス	ICRF_LSN

システム・カタログ

論理ファイル、ページおよびシステム・カタログの構成

【特許請求の範囲】

【請求項1】 ログと、プロセッサと、記憶されたページが前記プロセッサとの間でステージングされる記憶サブシステムとを備えた段階式コミット・トランザクション管理システム（TMS）において、ヘッダ・ページと、状況ページ（SMP）と、少なくとも1つのデータ・ページとを有する修正可能なページのファイルをアーカイブ保存するための方法であって、

前記プロセッサが各トランザクションにตอบสนองして、前記ページのうちの所定のページに対して選択的修正を実行し、ページ状態に対する各修正が前記ログに記録され、昇順ログ・シーケンス番号（page_LSN）が割り当てられ、前記page_LSNが前記ページ上に記録され、（a）前記ヘッダ・ページ上の第1の大域的ログ・シーケンス番号（ICBU_LSN）と、それぞれ前記各データ・ページに対応し、前記ページが直前のアーカイブ動作または複写動作以降に修正されたこと示す、前記SMP上の複数の状況ビット（ICB）とを生成し維持する段階と、（b）前記ICBU_LSNを最大値に設定し、第2の大域的ログ・シーケンス番号（ICRF_LSN）を設定し、システム・カタログ内に該ICRF_LSNを記録する段階と、（c）前記SMP内の前記ICBを走査することによってアーカイブ保存されるページのセットを確認し、別の形でセットされたサブセットをリセットする段階と、（d）段階（c）で走査されリセットされたサブセットの各ページを順にラッチし、複写し、ラッチ解除することにより、増分式複写のために、完全複写に関するICBの状態に関係なくすべてのページについて複写動作を実行する段階と、（e）前記ICBU_LSNを現end_of_log_LSNの値に設定し、該ICBU_LSNを前記ヘッダ・ページに記録する段階と、（f）各ページ修正に応じて、前記page_LSNが前記ICBU_LSNよりも小さい場合に前記SMP内に対応ICBを設定し、該設定動作をREDO_onlyログ・レコードによって前記ログ内に記録する段階とを含むことを特徴とする方法。

【請求項2】 前記ファイル内の各ページに、他の標像から独立した昇順のページ番号が記録されており、前記システム・カタログが、前記ファイルの各完全コピーまたは増分式コピーに関する情報を記録するエントリを含み、各エントリが、前記記憶サブシステム内の前記ファイルの各アーカイブ・コピーの位置を指すポイントと、請求項1の段階（e）によるend_of_log_LSNにセットされたICRF_LSNとを含み、前記方法が、

（g）ログすることなしにデータが前記ファイルにロードされるたびに各ファイルの完全コピーを作成し、該完全コピーが作成されるまでデータ・ページの修正を中断し、前記システム・カタログ内に前記ポイント位置とコピー・タイプ（完全）とICRF_LSNとを記録する段階と、（h）前記システム・カタログへのエントリを

含む各ファイルの増分式コピーまたは完全コピーを予定に従ってまたは都合に合わせて作成する段階と、（i）前記記憶サブシステム内の前記ファイルが使用できない状態の場合、前記システム・カタログによって定義されるような完全コピーおよび増分式コピーの内容をページ番号に従ってマージすることによって前記ファイルを回復し、マージされたページ・セットを、前記ICRF_LSNよりも大きいまたは等しいLSNを有する前記ログ・レコードに従って修正する段階と、

10 を含むことを特徴とする、請求項1に記載の方法。

【請求項3】 前記ログ内に記録された動作が、REDO/UNDOタイプのものとREDO_onlyログ・レコードタイプのものとを含み、前記ヘッダ・ページに対するICBU_LSN修正が、REDO_onlyログ・レコードを介して前記ログ内に記録されることを特徴とする、請求項1に記載の方法。

【請求項4】 前記TMSが、すでに修正済みのまたは現在修正中のページのリストを保持し、前記方法の諸段階が、（a'）ICBU_LSNを最大値に設定し、修正されたリスト内に各ページをSラッチする段階と、

20 （b'）前記ICBU_LSNを非最大値に設定し、ICRF_LSNの値を選択する段階と、（c'）前記SMP内のICBを走査することによってアーカイブ保存されるページのセットを確認し、別の形でセットされたサブセットをリセットする段階と、（d'）段階（c）で走査されリセットされたサブセットの各ページを順にラッチし、複写し、ラッチ解除することにより、増分式複写のために、完全複写に関するICB状態に関係なくすべてのページについて複写動作を実行する段階とを含むように修正されることを特徴とする、請求項1に記載の方法。

【請求項5】 ログと、プロセッサと、記憶されたページが前記プロセッサとの間でステージングされる記憶サブシステムとを備えた段階式コミット・トランザクション管理システム（TMS）において、修正可能なページのデータベースをアーカイブ保存するための方法であって、

前記プロセッサが各トランザクションにตอบสนองして、ページのうちの所定のページに対して選択的修正を実行し、各ページ状態に対する各修正が前記ログに記録され、昇順ログ・シーケンス番号（page_LSN）が割り当てられ、前記ページLSNが前記ページ上に記録され、

（a）（1）ヘッダ・ページと、少なくとも1つのデータ・ページと、前の複写動作の実行以降に前記データ・ページが修正されたかどうかを示す各データ・ページに関する状況ビット（ICB）を含む、少なくとも1つのスペース・マップ・ページ（SMP）とを備える区画に関する、データベース制御ブロック（DBCB）を呼び出し、（2）前記DBCB内の第1の種類のLSN（ICBU_LSN）を最大値に原子的に設定し、第2の種

類のLSN (ICRF_LSN) を現end_of_log LSNによって設定された値に設定し、システム・カタログ内に前記ICRF_LSNを記録し、(3)最終複写動作以降に更新されたことを示すデータ・ページのすべてのICBをリセットすることによって各SMPについて複写されるデータ・ページのセットを確認し、前記ログ内にリセットされたすべてのICBを記録することによって、前記データベースの前記区画に対する複写動作を初期設定する段階と、

(b) (1)各SMPについて、段階(a) (3)で識別された各データ・ページをラッチし、複写し、ラッチ解除し、(2)前記DBCBおよび前記ヘッダ・ページ内の前記ICBU_LSNを現end_of_log LSN値に更新し、(3)動作をコミットすることによって、前記複写動作を実行する段階と、

(c)各ページ修正に応じて、前記page_LSNが前記ICBU_LSNよりも小さい場合に状況ページ内に対応ICBを設定し、該設定動作を前記ログ内に記録する段階とを含むことを特徴とする方法。

【請求項6】ログと、プロセッサと、記憶されたページが前記プロセッサとの間でステージングされる記憶サブシステムとを備えた段階式コミット・トランザクション管理システム(TMS)において、ヘッダ・ページと、状況ページ(SMP)と、少なくとも1つのデータ・ページとを有する修正可能なページのファイルをアーカイブ保存するための方法であって、前記ファイル内の各ページに他の標徴から独立して昇順ページ番号が記録されており、

前記プロセッサが各トランザクションに回答して、前記ページのうちの所定のページに対して選択的修正を実行し、ページ状態に対する各修正が前記ログに記録され、昇順ログ・シーケンス番号(page_LSN)が割り当てられ、前記page_LSNが前記ページ上に記録され、(a)前記ヘッダ・ページ上の第1の大域的ログ・シーケンス番号(ICBU_LSN)と、それぞれ各データ・ページに対応し、前記ページが直前のアーカイブ動作または複写動作以降に修正されたことを示す、前記SMP上の複数の状況ビット(ICB)とを生成し維持する段階と、(b)前記ICBU_LSNを最大値に設定し、第2の大域的ログ・シーケンス番号(ICRF_LSN)を設定し、前記ファイルの各完全コピーまたは増分式コピーに関する情報を記録するエントリを含み、各エントリが、前記記憶サブシステム内の前記ファイルの各アーカイブ・コピーの位置を指すポインタと、end_of_log LSNに設定されたICRF_LSNとを含む、システム・カタログ内に前記ICRF_LSNを記録する段階と、(c)前記SMP内の前記ICBを走査することによってアーカイブ保存されるページのセットを確認し、別の形でセットされたサブセットをリセットする段階と、(d)段階(c)で走査されリセットさ

れたサブセットの各ページを順にラッチし、複写し、ラッチ解除することにより、増分式複写のために、完全複写に関するICBの状態に関係なくすべてのページについて複写動作を実行する段階と、(e)前記ICBU_LSNを現end_of_log LSN値に設定し、該ICBU_LSNを前記ヘッダ・ページに記録する段階と、

(f)各ページ修正に応じて、前記page_LSNが前記ICBU_LSNよりも小さい場合に前記SMP内に対応ICBを設定し、該設定動作をREDO_onlyログ・レコードによって前記ログ内に記録する段階と、(g)ログすることなしにデータが前記ファイルにロードされるたびに各ファイルの完全コピーを作成し、該完全コピーが作成されるまでデータ・ページの修正を中断し、前記システム・カタログ内に前記ポインタ位置とコピー・タイプ(完全)とICRF_LSNとを記録する段階と、(h)前記システム・カタログへのエントリを含む各ファイルの増分式コピーまたは完全コピーを予定に従ってまたは都合に合わせて作成する段階と、

(i)前記記憶サブシステム内の前記ファイルが使用できない状態の場合、前記システム・カタログによって定義されるような完全コピーおよび増分式コピーの内容をページ番号に従ってマージすることによって前記ファイルを回復し、マージされたページ・セットを、前記ICRF_LSNよりも大きいまたは等しいLSNを有する前記ログ・レコードに従って修正する段階と、を含む方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、複写によってアプリケーションの実行が妨げられない、完全なまたは増分式のバックアップ複写による、情報操作システムにおけるデータの保存に関する。より詳細には、本発明はログ・ベースの段階式コミット・トランザクション管理システム(TMS)において、修正可能なページのファイルをアーカイブ保存するための方法に関する。

【0002】

【従来の技術】以下の説明は、多段階ログ・ベース・トランザクション管理システムの諸態様、およびそうしたシステムにおいて修正されたデータ・ページをアーカイブ保存することを対象とする。

【0003】

トランザクション管理システムの諸態様 TMSは一般に、CPUと、修正可能なページのデータベースを記憶する外部DASDサブシステムと、ページにアクセスしそれを記憶装置とCPUの間で転送するための手段を含む。CPUで実行されるアプリケーションは、「トランザクション」を通じてページと対話する。「トランザクション」とは、回復可能な資源の一貫状態を、全ての中間点で必ずしも一貫性を保存することなく、別の一貫状態に変形させる一連の動作を含む、論理作業単位と定義される。すなわち、トランザクションと

は、データベース内の選択されたページをある状態から別の状態に原子的に変化させる有界の一連の動作である。したがって各トランザクションは、完了するか、または打ち切らなければならない、中間状態は許されない。

【0004】ログ・ベースのTMSでは、ページに対するすべての変更は、REDOレコードおよびUNDOレコードの形でログに書き込まれる。物理的ログは、DASDの予約部分やテープ・サブシステムなどでよい。ここで、REDOレコードまたはUNDOレコード自体が、変更または修正されたページをそのDASD記憶位置に書き戻す前に、ログに書き込まれる。これをライト・アヘッド・ロギングと言う。これらの変更レコードは、完了点まで進行したトランザクションを再作成またはREDOするために、あるいは完了点まで進行していなかったトランザクションをロールバックまたはUNDOするために使用される。

【0005】TMSは、各トランザクションが、BEGINプリミティブと、COMMITプリミティブまたはABORTプリミティブ、およびENDプリミティブによって境界を区切られることを特徴とする。すなわち、トランザクションは次の形である。

"BEGIN ops, ops, ..., ops COMMIT/ABORT ..END"

【0006】マルチタスク処理が可能なシステムなどでは、あるトランザクションは開始されたばかりで、他のものは「飛行中」で、さらに他のものはコミットされまたは打ち切られており、さらに他のものはついさっき終わったばかりというように、複数のトランザクションが完了時間が重なり合う。これは、トランザクションがデータベース内の選択されたページに対して何らかの更新を実行し、トランザクションがその正常終了(END)または一貫性中間点(COMMIT/ABORT)に到達する以前に障害が発生するとき、それらの更新がUNDOされることを段階式コミット・システムが保証することを意味する。もしそうでなければ、正常終了(END)または一貫性中間点(COMMITまたはABORT)に到達したトランザクションはREDOされる(COMMITまたはABORTを含む期間中でENDの直前に)。

【0007】回復動作を遂行するためには、修正されたページと対応ログ・レコードとの間で情報の同時性を提供しなければならない。これは、修正された各ページに固有な昇順のログ・シーケンス番号(LSN)を含め、ログに書き込まれるREDO/UNDOレコードにページ番号を含めることによって達成される。したがって、修正されたページがログ内のその変更レコード位置をインデックスし、ログ・レコードが修正されたページをインデックスする。

【0008】これらの態様の詳細な実施例は、米国特許第5043866号、および第4507751号明細書に出ている。

【0009】米国特許第5043866号明細書は、TMSにおいて順方向または逆方向の回復中に使用しなければならないログの範囲を定義する1対のログ・シーケンス番号が、周期的に決定されチェックポイント情報としてログ内に記憶される場合、回復時に、チェックポイントを検索し、ログの分析パスを回避するために回復アルゴリズム中で用いられるその構成要素間の機能比較を行うことを教示している。

【0010】米国特許第4507751号明細書は、同期点(BEGIN、COMMIT、ABORT)の発生時に、バッファ内容を第2のログに書き出すことによって対処される、揮発性メモリまたはバッファ内に記憶されたデータのジャーナル処理(ロギング)を開示している。障害が発生する場合、第2のログのロールバック(REDO/UNDO)処理によってバッファの前の状態が再確立され、そうでない場合は第1のログが使用される。

【0011】アーカイブ保存

当技術分野では、アーカイブ保存とは、冗長性と機密保護の現情報に関して不定期にアクセスされるシステムの一部分にページを複写するという、体系的な記憶管理の慣行を指す。アーカイブ保存またはバックアップ・コピーは経時的に完全または増分式に行われ、かつ予定に従ってまたは都合に合わせて行われる。明らかに、増分式複写は完全複写よりも消費資源が少ない。

【0012】IBMテクニカル・ディスクロージャ・ブルテン、No. 25、pp. 3730~3732(1982年12月発行)に所載のクラス(Crus)他の論文"Incremental Database Log Image Copy"は、スペース・マップ・ページ(SMP)と名付けられた共通の状況構造を走査し、最後の複写動作以降にページが変更されたことを示す状況ビット(増分式状況変化ビット(ICB)と呼ぶ)を有するデータ・ページだけを複写することによって、データベース内で修正されたページを増分式に複写する方法を開示している。より詳細には、複写動作の後でデータ・ページが最初に修正されるとき、データ・ページ中およびスペース・マップ・ページ中でICBがセットされる。データ・ページが修正されており、そのICBがすでにセットされている場合は(最後の複写動作以降の修正を示す)、SMPのICBを修正するために何も行われない。ICBをそのページに置く理由は、そうすると、データ・ページの更新ごとにSMPを訪問するというオーバーヘッドが減少するからである。複写の目的で、SMPが走査され、そのICBがセットされている各ページが複写され、次いでデータ・ページおよびSMP上の対応ICBがリセットされる。

【0013】IBM関係データベース・システムDB2では、修正されたページのディスクへの書き込みは、「即時コミット」を有効にするため、ICBビットをリセットした直後に実行される。これは、DASD書き込み動作

のバッチ処理ではなく、各ページごとに個別に実行されるので、かなりのオーバーヘッド・コストになる。

【0014】

【発明が解決しようとする課題】本発明の目的は、ログ・ベースの段階式コミットTMSのデータベース中の修正されたページの増分式複写に伴うオーバーヘッドを減少させるための方法および手段を提供することである。

【0015】本発明の他の目的は、アプリケーションの実行を妨げないログ・ベースの段階式コミットTMS内で修正可能なページのデータベースをアーカイブ保存するための方法および手段を提供することである。

【0016】本発明のさらに他の目的は、修正可能なページのデータベースへのアクセスを共有する複数ログ・ベース段階式コミット・トランザクション管理システムで動作可能な方法および手段を提供することである。

【0017】

【課題を解決するための手段】本発明は、ログ・ベースの段階式コミットTMSの一部として使用可能な構造を利用する。これは、スペース・マップ・ページ(SMP)、ページ更新状況ビット(ICB)、ヘッダ・ページ、データベース制御ブロック(DBCB)、ログ、2種類のLSN(ICBU_LSNとICRF_LSN)、およびシステム・カタログを含む。SMPは各データ・ページに関するICBを含み、ヘッダ・ページはデータベースに関するICBU_LSNを格納し、DBCBは仮想記憶域中のICBU_LSNを維持し、システム・カタログはデータベースに関するICRF_LSNを格納する。

【0018】各SMPは、最後の複写動作以降にi番目のページが更新されたかどうかを示す状況ビット(ICB)を含む、複数のデータ・ページに関する割当て状況を追跡する。ICBは、複写動作が実行された後で最初にページが更新される時にセットされ、前記動作によってページが複写される時にリセットされる。「ヘッダ・ページ」は、データベースの制御または管理構造である。

【0019】DBCBは、データベースがオープンされる時に仮想記憶域中で生成され維持される。これはデータベースの現状態を反映する主要制御構造であり、本発明の方法における初期設定段階の一部として含まれる。またこれは、2つのLSNのうち的一方(ICBU_LSN)のためのレポジトリでもある。第1のLSNは、「イメージ複写ビット更新LSN」(ICBU_LSN)と呼ばれる。第2のLSNは、「イメージ複写順方向ロールLSN」(ICRF_LSN)と呼ばれる。

【0020】ICBU_LSNの値により、SMP内のどれか所与のデータ・ページに関するICBをセットする必要があるかどうかが決まる。ICBU_LSNの初期値は、データベースがDASD記憶装置に最初にロードされた後の、end_of_log LSNに設定される。

【0021】ICRF_LSNは、最後の完全コピーまたはその後の増分式コピーを再ロードした後でデータベースを回復するためにREDO(REDO動作)しなければならないログ・レコードを識別するために、ログをそこから走査しなければならないLSNを示す。

【0022】本発明の方法により、ICBU_LSNを2段階で使用することにより、従来の問題点を解決することができる。

【0023】第1段階では、ICBU_LSNが最大値にセットされ、SMP中のICBを調べることによって複写順方向ロールLSN(ICRF_LSN)とページのコピー・セットの両方が確認される。増分式複写では、コピー・セットは、そのICBが最後の複写動作以降に修正されたことを示すように変更された、ページのセットである。

【0024】第2段階では、コピー・セット中のページを複写した後で、ICBU_LSNを現end_of_log LSNにセットし、ヘッダ・ページに記録する必要がある。複写では、ページが短時間に逐次化されて、複写動作の実行と同様にページの修正とログが行えるようになる。

【0025】

【実施例】

発明の方法を実行するためのCPU環境

本発明は、システム内の各CPUが、IBMのMVSオペレーティング・システムを有するIBM/360またはIBM/370で構成されたCPUタイプの構成中で好都合に実行できる。IBM/360構成のCPUは、米国特許第3400371号に十分に記載されている。外部記憶装置へのCPUの共用アクセスを含む構成は、米国特許第4207609号明細書に記載されている。

【0026】また、MVSオペレーティング・システムは、IBM刊行物GC28-1150、"MVS/Extended Architecture System Programming Library: System Macros and Facilities", Vol. 1に記載されている。ローカル・ロック管理、割込みまたはモニタによるサブシステム呼出し、タスクの記入および待機など、標準のMVSまたは他のオペレーティング・システムのサービスの詳細は、省略する。これらのOS処理は、当業者にはよく理解されるであろう。

【0027】

本発明の実施態様およびトランザクション管理システム
本発明の実施例の動作を理解するには、TMSが、所与のオペレーティング・システムのもとで実行されるアプリケーションとして表されることを想起されたい。この場合、アプリケーションは、たとえば、米国特許第4498145号に記載されているようなDB2型の関係データベース・システムのアプリケーションである。

【0028】オペレーティング・システムは、その1つのタスクとして、内部記憶域と外部記憶域を資源として

含むメモリを編成する。アプリケーションから見ると、メモリと記憶域は仮想式のものであり、実要求ページングLRU階層形式の記憶域によってバックアップされる。

【0029】ここで図1を参照すると、TMSがデータベース・システムの形で示されている。トランザクション・プロセス12は、その実行が複数のトランザクションの並行な重なり合った実行をサポートする、アプリケーション・プログラムでもよい。プロセス12によって実行されるトランザクションは、データベース管理システム・プログラム14 (DBMS) と入出力 (I/O) サービスを行うオペレーティング・システム16とを介して、データベース13にアクセスすることができる。データ・バッファ・プール18は、DASDに記憶されたデータベース13に関するデータ用に、CPUの主メモリからDBMS14に割り当てられる。またDBMS14には、システム・ログ21用のログ・レコードの記憶用に、主メモリからログ・バッファ20が割り当てられる。

【0030】DBMS14は、レコード・マネージャ22、バッファ・マネージャ23、ログ・マネージャ24、回復マネージャ25、および並行性マネージャ26を含む。

【0031】レコード・マネージャ22は、データベース13のデータ構造および記憶空間を調整する。それによって、トランザクションへのレコード・レベルのアクセス、およびデータベースのロード、複写、回復などシステム・ユーティリティへのページ・レベルのアクセスが提供される。バッファ・マネージャは、データベース13とバッファ・プール18の間でページを移動する。ラッチ・マネージャは、バッファ・プール18中の読み取られているまたは修正されているページに対する短期間の逐次化 (共用または専用) を提供する。ログ・マネージャ24はログ・レコードを生成し、それらをログ・バッファ20内で番号付き順序で組み立て、ログ21に書き込む。回復マネージャ25は、トランザクション・レベルの回復をサポートするためにログ・レコードを利用してデータベースを戻し、同時に並行性マネージャ26はロック・テーブル30を介してロックを実施する。

【0032】ここで図2を参照すると、チェックポイントおよびシステム障害に関するトランザクション・プリミティブ間での発生関係を時間で表している。障害の発生とそれ故の再始動の際に、回復マネージャは再始動ファイルまたはその等価物から最新のチェックポイント・レコードのアドレスを獲得し、システム・ログ内でチェックポイント・レコードを探し出し、ログをその点から終りまで順方向に探索を続ける。このプロセスの結果として、資源を一貫状態に復元するために、回復マネージャは、UNDOする必要のあるトランザクションとREDOする必要のあるトランザクションの両方を決定す

ることができる。

【0033】各トランザクションは、5つのクラスのうちの一つに類別される。たとえば、タイプT1のトランザクションは、チェックポイント時間Tcより前に完了した。タイプT2のトランザクションは、時間Tcより前に開始し、時間Tcより後でシステム障害時間Tfより前に完了した。またタイプT3のトランザクションは、時間Tcより前に開始したが、時間Tfより前には完了しなかった。タイプT4のトランザクションは、時間Tcより後で始動したが、時間Tfより前に完了された。最後に、タイプT5のトランザクションは、時間Tcより後に開始したが、時間Tfまでには完了しなかった。チェックポイント時間に、修正されたページがすべてDASDに書き込まれていたと仮定すると、トランザクションT2およびT4はREDOを施され、トランザクションT3およびT5はUNDOを施される。

【0034】回復マネージャは、2つのリストを始動する。UNDOリストは最初、チェックポイント・レコードにリストされたすべてのトランザクションを含み、一方REDOリストは最初空である。回復マネージャは、チェックポイント・レコードから始めてログを順方向に探索する。回復マネージャが所与のトランザクションに関するBEGINトランザクション・レコードを見つけた場合、そのトランザクションをUNDOリストに加える。同様に、回復マネージャが所与のトランザクションに関するCOMMITレコードを見つけた場合は、そのトランザクションをUNDOリストからREDOリストに移す。

【0035】回復マネージャは、ログ中を順方向に動作してすべてのトランザクションを再実行し、逆方向に動作してUNDOリスト内のトランザクションを取り消す。

【0036】

好ましい実施例の方法およびプロトコル・レベルの詳細最初に、データ・ベース中および仮想記憶域中のデータ構造と、ロギング・プロトコルおよびラッチング・プロトコルについて述べる。第2に、複写動作に関係する処理と、(必要ならば) ICBをセットするためのトランザクションの更新と、バックアップ・コピーを使用するデータ・ベースの回復について述べる。第3に、その実行が障害によって割り込まれる場合にコピーのロールバックに使用される処理について説明する。第4に、DB2のようなシステムに関連するバッファ・プールを介してではなく、DASD記憶域からデータを直接複写することが可能な場合の本発明の方法の使用の仕方について述べる。最後に、共用DASD記憶域を有する多重システム・トランザクション処理システムへの本発明の方法の拡張について述べる。

【0037】データ・ベースのデータ構造

本発明の目的では、データ・ベースはデータ・ページお

よび補助構成を含む。この点に関して、ページのファイルはデータ・ベースと機能的に同等と見なす。補助構成は、割振り状況などを追跡するために、データ・ベース内にページを含む。ユーザ・データを含むページは「データ・ページ」と呼び、データ・ページの割振りおよび空間可用性状況を追跡するシステム所有のページは「スペース・マップ・ページ」(SMP)と呼び、情報に係るシステムを含むシステム所有のページは「ヘッダ・ページ」と呼ぶ。

【0038】DB2の場合と同様に、大きなテーブルを、それぞれ別々のオペレーティング・システム・ファイルである多数の区画に分割することができる。各区画はヘッダ・ページと、1つまたは複数のスペース・マップ・ページと、非常に多くのデータ・ページを有することになる。アーカイブ・コピーには、データ・ページのコピーだけではなく、SMPおよびヘッダ・ページのコピーも含まれる。

【0039】SMPはいくつかのデータ・ページの割振り状況を追跡する。またSMPは、これがカバーする各データ・ページごとに1つのICBを含む。ICBの目的は、最後に複写動作が実行されて以降に、対応するデータ・ページが修正されたかどうかを追跡することである。

【0040】ヘッダ・ページ内では、"image_copy_bit_update_LSN"(ICBU_LSN)と呼ばれるフィールドが維持される。1つのファイルにはICBU_LSNが1つだけある。ICBU_LSNの目的は、それを使用することによって、データ・ページを更新するトランザクションが、そのデータ・ページに関するSMP内のICBをセットする必要があるかどうかを効率的に判定できるようにすることである。これが望ましいのは、ICBは、それが最後の複写動作以降のデータ・ページへの最初の更新である場合にだけセットする必要があるからである。変更を実行する必要があると判定するだけでも、DASDの入出力、バッファ・プール内のSMPの位置決定と固定、そのラッチ、適切なICBの探索、「その固定解除」などのオーバーヘッドが含まれるので、データ・ページを後で更新するためにSMPにアクセスすることは避けなければならない。データ・ベースが最初にロードされた後、データ・ベースへの更新が許される前に、データ・ベースの完全コピーを取らなければならない。これはログすることなしにローディングが行われたと仮定している。これは、媒体回復が始まる前に繰り返されなければならないロードなしに、媒体回復が可能でなければならないときに必要である。完全複写動作では、ICBU_LSNの初期値を、データベースを複写した後の現end_of_log_LSNとして確立する。

【0041】各アーカイブ・コピーに複写順方向ローカルLSN(ICRF_LSN)が関連している。媒体回復中、これは、関連アーカイブ・コピー(最近の完全コピ

ーおよびその後の増分式コピー)を再ロードした後でデータ・ベースを回復するためにその更新を再実行しなければならないログ・レコードを識別するためにログをそこから走査しなければならないLSNである。ICRF_LSNは、アーカイブ・コピーを含むファイル名や完全コピーと増分式コピーのどちらが取られたかなどの情報と共にシステム・カタログ中に記憶される。

【0042】仮想記憶域内のデータ構造

データ・ベースが「オープン」状態にある限り、DB2のようなシステムは仮想記憶域内にそのための「データ・ベース制御ブロック」(DBCB)を維持する。

【0043】更新トランザクションによって、ICBU_LSNをルックアップする動作を効率的にするために、データ・ベースがオープンになるときに、ICBU_LSNの値が、ヘッダ・ページからDBCB内のフィールドに複写される。その後、このフィールドが、増分式コピーの方法および手段によってのみ更新される。

【0044】TMSは、回復のためにその更新を記述するログを使用する。ページが更新されると、システムは、更新を記述するログ・レコードに、システム規模の単調に増加するログ・シーケンス番号(LSN)を割り当て、そのLSNを"page_LSN"と呼ばれるフィールド内の修正されたページに記録する。

【0045】複写動作は、単一のトランザクションとして実行され、ICBをリセットするとき、UNDO-REDOログ・レコードを書き込むものと仮定する。

【0046】ページ・ラッチ

TMSは、単一のページ上で並行した読取り動作と更新動作または多数の並行した更新動作を逐次化するためにページ・ラッチをサポートする。共用(S)ラッチは読取り動作によって使用され、専用(X)ラッチは更新動作によって使用され、SモードとXモードの間では通常の互換性規則が適用される。ページ・ラッチは、ページのファジー・コピーを行うために、複写動作によって使用される。すなわち、複写ページがコミットされていないデータを有していてもよい。

【0047】正確判定基準

増分式コピー(IC)に関する正確さの要件は、最新の完全コピー(FC)あるいはそのFCの後でかつ現ICの前に行われたICにおいて、その効果がすでに捕捉されている更新を除き、ログ中の現ICのICRF_LSN点より前に書き込まれたログ・レコードによって表されるすべてのデータ・ページ更新の効果を、現ICが捕捉しなければならないことである。

【0048】FCに関する正確さ要件は、ログ中の現FCのICRF_LSN点より前に書き込まれたログ・レコードによって表されるすべてのデータ・ページ更新の効果を、現FCが捕捉しなければならないことである。

【0049】バッファ・プール、システム・カタログ、および入出力の概念

バッファ・プールとは、計算または制御動作で中間結果、データ、または値を一時的に記憶するために留保されているアドレス可能な主メモリの一部分である。本発明では、他の場所で読み取り、修正し、複写する目的で、ページに関する入出力がバッファ・プール内に複写される。

【0050】システム・カタログとは、ファイルの記憶位置に関するディレクトリである。これはまた、状況情報を含むこともできる。本発明では、これは、ファイル名、完全コピーか増分式コピーかの区別、日時、記憶される装置、および関連ICRF_LSNを含む。

【0051】入出力は、ページにアクセスするための付随的な活動を表すために使用される用語である。

【0052】単一ランザクション管理システムにおけるアーカイブ複写

以下の諸段では、単一のTMS環境で使用する複写動作を機能的に記述する。この場合、複写動作ではデータベース・ページがTMSバッファ・プールに入れられる。その後、データがDASDから直接読み取られるとき、すなわちページがTMSバッファ・プールに入れられないときに使用される複写動作について説明する。次の節では、多重システムの共用ディスク環境においてTMSのために使用される複写動作について説明する。

【0053】呼出し可能なソフトウェア機能として表される複写方法、すなわち複写動作

この節では、(a) 複写動作、(b) トランザクションによるデータ・ページの更新、および(c) 媒体障害後のデータ・ベースの回復を扱う。障害によって複写動作の実行が中断される場合のロールバック処理について述べる。

【0054】アーカイブ複写は次の諸段階を含む。

【0055】同時に同じデータ・ベースに対して複写動作が1つだけ実行されるようにするために、システム・カタログ中に所与のデータ・ベースに関する複写動作の初期設定を登録する。

【0056】データ・ベースがまだオープンしていない場合に、データ・ベースをオープンし、データ・ベースのヘッダ・ページを読み取って、DBC B中のICBU_LSNの値をセットする。これによってDBC Bが生成される。データ・ベースがすでにオープンされていた場合は、DBC B内のICBU_LSNはすでにセットされている。

【0057】このデータ・ベースのすべてのSMPが、バッファ・プール内に確実に存在するようにする。まだバッファ・プール内にはないSMPについては、当然、それらのSMPをバッファ・プールに送るために入出力を開始し、これらの入出力が完了するのを待つ必要がある。

【0058】このSMPの事前取り出しは、DBC B中のICBU_LSNフィールドが可能な最大値(X'F

F....FF'、次の段階を参照)を有する期間を最小にするための最適化である。

【0059】DBC B中のICBU_LSNの値を原子的にX'FF....FF' (すなわち可能最大値)に変更する。

【0060】媒体回復中に使用するために、適切なシステム・カタログ中に現end_of_logのLSNをICRF_LSNとして記録し、その動作をログする。

【0061】データ・ベース内の各SMPについて、以下のよう初期設定を行う。

・SMPをX (専用) ラッチする。

・SMP内のすべてのICBを検査して、“1”の値を有するICBを、“0”にリセットする。ICでは、対応するデータ・ページは、最後の複写動作(完全または増分)以降に修正されているはずのものである。完全複写の場合は、すべてのデータ・ページが複写される。

・リセットされたすべてのICBを記述するREDO/UNDOログ・レコードを、このSMP上に書き込む。ICでは、どのデータ・ページを複写するかを知るために後で使用できるように、ログ・レコードのコピーを仮想記憶域に保持する。

・SMPのラッチを解除する。

【0062】ICRF_LSNの値が選択された後で、新しいSMPがデータ・ベースに追加された場合、それらはこの複写動作の一部として検査する必要はないことに留意されたい。

【0063】そのような各ICB_iについて、以下のことを行う。

【0064】i番目のデータ・ページをS (共用) ラッチする。

【0065】i番目のデータ・ページをアーカイブ・コピー・ファイルに複写する。

【0066】i番目のデータ・ページをラッチ解除する。

【0067】FCについては、このSMPでカバーされるすべてのデータ・ページが複写される。

【0068】これらの動作を極めて効率的にし、かつ入出力のコストおよび遅延を減少させるために、バッチ式入出力(すなわち、1回の入出力呼出しで1ページよりも多く読む)および所望のデータ・ページの事前取出しを使用することができる。また並列処理を用いて、また複写動作の経過時間をさらに減少させることもできる。

【0069】関係する動作には以下のものが含まれる。

・現end_of_log LSNを得る。

・DBC Bおよびヘッダ・ページ内のICBU_LSNフィールドを、現end_of_log LSNの値に更新する。

・REDO_onlyログ・レコードを使用して、その更新をヘッダ・ページにログする。

【0070】アーカイブ複写ランザクションをロールバックすべき場合も、古いICBU_LSNに戻るの

はなく、新しいICBU_LSNを保持するのが正しいはずである。これが、REDO/UNDOログ・レコードではなく、REDO_onlyログ・レコードを使用する理由である。これに関連して、この方法は、共用DASD環境でヘッダ・ページ更新のUNDOの何らかの特別な扱いを回避する。

【0071】データ・ベース内のそれぞれのSMPについて、以下のことを行う。

【0072】ICについて、そのSMPに関して前に書き込まれたログ・レコードの、キャッシュされたバージョンを検査することによって、その動作によって（“0”に）リセットされたICBを決定する（ICB_iはi番目のデータ・ページと関係する）。

【0073】複写動作の終了時に（すなわち、すべての関連データ・ページが複写された後で）、次のことを行う。

【0074】この複写動作が完了しており、したがってこのイメージ複写トランザクションがコミットされた後は、このデータ・ベースに関するイメージ複写動作のその後の呼出しが遅延なしで実行を開始できることに留意されたい。

【0075】複写トランザクションをコミットする。

【0076】複写動作の実行中に並行更新が許されない場合は、その方法は上記の方法の一部分である。ICRF_LSNの値は、必要なページをすべて複写した後で選択される。この複写動作には、ページのラッチは必要でない。

【0077】ページを更新するトランザクションのための方法または論理

更新トランザクションは、データ・ページを更新する間に、そのデータ・ページに関するICBをSMP中で（“1”に）セットする必要があるかどうかを判定しなければならない。これは、（この更新を行う前に）更新されるページの現page_LSNをICBU_LSNと比較することによって行われる。この検査により、データ・ページが複写動作の最後の実行以降にすでに修正されているか、それともその時以降で最初の更新であるかを判定する。

【0078】更新トランザクションが従う方法または論理は、次のとおりである。

【0079】データ・ページをXラッチする。

【0080】現page_LSNがDBC内ICBU_LSN（更新前）よりも小さいときは、次のことを行う。

【0081】このデータ・ページをカバーしSラッチするSMPにアクセスする。

【0082】複写動作が実行を開始したすぐ後にICBU_LSNがX'FF...FF'にセットされるので、このデータ・ベースに対するすべての更新がこの動作の実行中に関連SMPを訪問することになる。

【0083】ICBがまだセットされていない場合は、それをセットし、REDO_onlyログ・レコードを使用する動作をログする。SMP上ではSラッチしか獲得されていないので、複数のトランザクションがSMPを並行して更新中であることがあり得る。Sラッチ処理は、そのようなホット・スポット・ページ上での並行性を改善するために行われる。その結果、page_LSNフィールドのみならずICBに対する更新も慎重に行わなければならない（たとえば、page_LSNフィールドが別のトランザクションによって高い値に更新されている場合はそれを更新してはならず、比較交換の方法または論理を使用すべきである）。

【0084】次の場合には、ICBはすでにセットされていることがあり得る。

【0085】DBC内ICBU_LSNがX'FF...FF'にセットされた状態で複写動作が実行中に、データ・ページが1回または複数回更新された。その間にSMPを何度も訪問することを回避する1つの方法は、ICBU_LSNがX'FF...FF'の間に、データ・ページのICBがすでにセットされたことを示すことのできるフラグを、そのページに関するバッファ・マネージャのページ制御ブロック（PCB）内に立てることである。複写動作は、データ・ページ上にラッチを保持している間、そのフラグをリセットする。そのフラグは、そのデータ・ページがある更新から別の更新にキャッシュされたままである場合、すなわちそのページに関するバッファ・スロットがスチールされていない場合だけ役立つ。ページが置き換えられ再びキャッシュされた場合、PCBフラグがICBがセットされていないことを意味するように初期設定される。

【0086】SMP更新のためにREDO専用レコードとしてログ・レコードを書く理由は、そのデータ・ページを更新したトランザクションがロールバックすべき場合でも、SMP内でICBの値が（“1”に）セットされたままではなければならないからである。これは、別のトランザクションによる同じデータ・ページへのその後の更新では、page_LSNが最初の更新によってICBU_LSNよりも大きい値にセットされているはずだから、ICBはセットされているはずだと仮定するからである（以下参照）。

【0087】SMPのラッチを解除する。

【0088】DPを更新し、更新をログし、DPのpage_LSNを、書き込まれたばかりのログ・レコードのLSNにセットする。

【0089】DPをラッチ解除する。

【0090】ICBU_LSNが最大値にセットされたままである時間を減少させるための最適化
バッファ・プール内にその時キャッシュされているページのリストを維持するDB2などのTMS内では、ICBU_LSNが最大値になっている時間を減少させるこ

とができる。これは以下のようにして実施される。

【0091】ICBU__LSNを最大値にセットした後、変更されたページのリスト上の各ページをSラッチする。

【0092】ICBU__LSNを現ログ終端（非最大値）にセットする

【0093】ページを複写する。

【0094】この最適化では、ICBU__LSNを非最大値にセットした後でページがアクセスされ複写されるので、ICBU__LSNが最大値にセットされている時間10が減少する。

【0095】ICBをセットする必要がある場合は、ICBがセットされICBセット動作がログされた後で、対応するDPの更新のロギングを行うことが重要であることに注意されたい。順序を逆にすると媒体回復が正しくなくなることがある。

【0096】新しく割り振られたページについては、ICBは常にセット（“1”）される。

【0097】複写動作実行中の障害

複写動作の実行が、何らかの障害によって中断される場合は、たとえば複写動作によって書き込まれたログ・レコードを使用する通常のロールバックの方法または論理により、その設定が複写動作によって修正されたICBの古い設定が復元される。他のICBの値が、ロールバック法によって、それらの元の値を含むように修正されない（すなわち、どのICBにもロールバック中に“0”が割り当てられない）ことが重要である。また、現複写実行に関するシステム・カタログ内の情報（たとえば、ICRF__LSNの値）は削除されることになる。

【0098】媒体障害後の回復

媒体回復のための方法または論理では、最後に完了した複写動作のICRF__LSNの値からログ走査を始める必要がある。この走査は、最新のFCおよび後続のすべてのIICから時間順にデータが複写された後で開始される。

【0099】実例

ここで図3を参照すると、本発明で使用される論理ファイルおよびページ編成が示されている。さらに詳細には、ファイルまたはデータ・ベースは、ヘッダ・ページ、スペース・マップ・ページ（SMP）、および複数のデータ・ページ（Dページ1～4）を含む。任意に「ファイルA」と名付けたデータが、CPU制御のロード動作の制御下で、DASD外部記憶装置内のデータベース13にロードされる。ログ・スペースとログ・オーバヘッドを節約するために、ロードはログを中断した状態で実行される。また、ロード動作の実行中、並行更新は許されない。

【0100】この動作の一部分として、各ページにページ番号が連続して割り当てられる。ヘッダ・ページの番号はページ0、SMPはページ1、Dページ1～4はベ

ージ2～5と番号が付けられる。ここでは、ページ番号は単にローカル・ページ・シーケンスを定義するものにすぎず、page__LSNと混同してはならない。後者は各ページに埋め込まれたポインタであり、ログされた後でDASD記憶装置に書き込まれたとき、最新のページ修正を含むログ中のREDO/UNDOレコードの位置を示す。最初に、page__LSNが0にセットされる。同様に、すべてのDページに関するSMP内のICBが0にセットされ、ヘッダ・ページ内のICBU__LSNも同様である。最後に、ファイルAに対するバックアップは存在しないので、システム・カタログ内にエントリはない。

【0101】図3から理解されるように、システム・カタログは、ファイルAおよび他のファイルの完全コピーまたは増分式コピーに関する位置ポインタと他のアーカイブ・コピーまたはバックアップ・コピー情報を幅広く含む。一般に、そのようなアーカイブ・コピーは、自動テープ・ライブラリなどの補助記憶装置に記憶される。バックアップ・コピーの回復使用には、テープ・ライブラリへのアクセスとそのテープ・ライブラリからDASD記憶装置へのステージングが必要である。

【0102】ここで図4を参照すると、初期ロードの後データ・ページに許される修正の前に、ファイルAの完全コピー（FC）を取った後のファイルAおよびシステム・カタログの状態が示されている。図4で、それがFC完了後の現end_of_logであったので、ICBU__LSNおよびICRF__LSNが100にセットされていることに留意されたい。ヘッダ・ページは、その後110のLSNを有するログ・レコードを書き込むことにより、ICBU__LSNのこの値を記録する方法で更新される。すなわち、ヘッダ・ページが、現ICBU__LSN値100を記憶しているログの位置がログ位置110である。

【0103】ここで、図5を参照すると、ファイルAのFCが1992年5月1日23時50分に取りられ、テープ・ライブラリ・アドレスT001に記憶された、システム・カタログ・エントリが示されている。FCが取られた後でDASD D001（図示せず）が失われる場合、ファイルAに関するDBMS回復動作は次の諸段階を含むことになる。

【0104】（a）システム・カタログからのファイルAの最近の完全コピーに関する現ログ終端（任意にEND__LSNと名付ける）とテープ位置を確認する。

【0105】（b）ファイルAのFCバックアップを、テープ位置T001からDASD D002（図示せず）にロードする。

【0106】（c）システム・カタログからのFCの後でファイルAの増分式コピー（IIC）を確認し、それに応じて更新する。（段階（c）のこの場合には、FCの後でファイルAのIICがまだ取られていない）

【0107】(d)それが最新のバックアップなので、システム・ログをICRF_LSN=100に位置決めし、page_LSN<ログ・レコードLSNの場合は、DASD D002にあるファイルAのページの更新に関するログ・レコードを適用する(この例では、ICRF_LSN=100からEND_LSNである)。

【0108】段階(d)中のログ走査の回数を減らすために、DBMSは修正されたシステム・カタログ内に配置された別のディレクトリを利用することもできる。この別のディレクトリは、更新のためにファイルAが開かれている間のログ・レンジからなることが好ましい。この例で「SYSLGRNG」と記されたこのディレクトリ(図示せず)は、ヘッダ・ページの更新に関する100から110のLSNを表す、ファイルAのエントリを有する。その後、回復動作が完了する。

【0109】ここで、図6ないし図10を参照すると、選択されたデータ・ページ(ページ1とページ3)のトランザクションによる修正が示されている。ヘッダ・ページがICBU_LSN=100とDページ1およびDページ3のpage_LSN=0を持つことを思い起こされたい。

【0110】Dページ1の更新では、図6に示す通り、ファイルAに関するICBU_LSN=100がまたデータ・ベース制御ブロック(DBCB)内に記録される。Dページ1のpage_LSN0はICBU_LSN100よりも小さいので、そのとき更新方法で、Dページ1を更新する前に、SMP内のICB1をセットしなければならない。ICB1の変更と、変更がログ(LSN212)内に記録されるLSNは両方とも、図7に示したようにSMP上に示される。次に、Dページ1の値がA1からB1に変えられ、LSN215に記録されたログになる。これらの変更は、図8の更新されたDページ1に示される。

【0111】Dページ3の更新では、Dページ3のpage_LSN0はICBU_LSN100よりも小さいので、図9からわかるように、SMP内のICB3はDページ3の更新前に1にセットされる。これに続いて、図10に示すように、Dページ3は値A3からC3に更新され、これがLSN326にログされたLSNもDページ3でマークされる。

【0112】ここで図11を参照すると、ファイルAが、更新されて、SMPおよびページ更新の変更を含むDASDに書き込まれたものとして示されている。

【0113】ここで図12ないし図23を参照すると、図11に描かれたファイルAのICが示されている。図12および図13は、DBCBとSMPによるIC動作に関するファイルAの初期状態を示す。最初の段階は、バッファ・プール内にSMPを読み込むことである。次に、図14に示すように、ICBU_LSNが16進表記で最大値「FFFF...FF」に変更される。これに

続いて、ICRF_LSNが、現end_of_log=463にセットされる。ここで、図15に示すように、システム・カタログが、ICが取られていること示す新しいエントリ、日時、テープ内のICの位置、およびICRF_LSNで更新される。

【0114】次に仮想記憶域内のSMPを更新することが望まれる。これは、SMPをラッチして、それを図16に示したように変更する必要があることを意味する。すなわち、SMPをラッチした後、「1」であるICBが「0」にリセットされ、図17に示したようなUNDOレコードがログに書き込まれ、その後にSMPがラッチ解除される。

【0115】IC複写動作は、そのICBが「1」から「0」にリセットされた、ヘッダ・ページ、SMPページおよびDページを複製する段階を含む。Dページを含む各ページを、ラッチし、テープT002に複写し、次いでラッチ解除する必要がある。図18に示すように、テープ上の各ページは、Dページ1のそれと同様に新しい。完璧を期して、図21に示すように、ICはDページ3のテープT002へのコピーを作成させる。最後の段階では、ICBU_LSNに関する現end_of_log値を獲得し、それをDBCBと更新されたヘッダ・ページの両方に複写し、それをヘッダ・ページに関してログし、図22および図23に示すヘッダ・ページ内にpage_LSNを置く。

【0116】さらに、再び図12ないし図23を参照して、図18に示すように、テープT002に複写された直後に、別のトランザクションがDページ1を更新しようとするものとする。図16に示すように、増分式に複写されるページのセットの決定の一環として、すべてのICBが「0」にセットされていることを想起されたい。したがって、Dページ1への現更新にはICB1をセットすることが必要である。ここで、Dページ1の場合、page_LSN=215である。これは、最大値にセットされたICBU_LSNよりも小さい。したがって、図19および図20に示すように、ICB1はSMP中で1にセットされ、Dページ1が更新され、両方のページがログされ、LSNが当該のページに記録される。

【0117】重要なことであるが、ICBU_LSNの値が100のままであった場合、Dページ1の更新によって、ICB1が「1」にセットされることにはならないはずである。したがって、後続のICは、値「C1」を有するDページ1を複写しないことになる。後続のICが完了すると、ICRF_LSNは504よりも大きくなる。次のIC後にファイルAが失われる場合、504よりも大きいICRF_LSNを使用したバックアップ・コピーの復元により、LSN504の更新が失われることになる。これにより、ICの間に新しい非最大値が確立されつつある間ICBU_LSNを最大値にセット

することの必要性が説明できる。コピー・セット中のすべてのページをラッチし、複写し、ラッチ解除することを含めてICを完了した後で、ICBU_LSNを非最大値にセットすることが必要であり、そうしないと、対応するICBをセットするために、あるいはICBがすでにセットされているかどうかを調べるために、ファイル内のページへのあらゆる更新でSMPへのアクセスが起こることになる。これは無駄だと考えられる。

【0118】この例では、Dページ3はまだ複写されておらず、したがって、ICBU_LSNの新しい非最大値は確立されていない。Dページ1を更新したトランザクションは、そのpage_LSNを比較するためにICBU_LSNの最大値を使用した。

【0119】ICBU_LSNの新しい非最大値をセットする際の他の要因

ICBU_LSNの新しい非最大値をセットする際に考慮すべき要因が他にもある。ここでは、比較のために古いICBU_LSNを使用し、したがってICBビットの設定を迂回した可能性のある更新トランザクションがなく、その新しいpage_LSNがICBU_LSNよりも大きいことが必要である。そのようなことは、更新トランザクションがそのpage_LSNをICBU_LSNの古い値と比較した後で、そのトランザクションがオペレーティング・システムによって待機中の活動または予定の活動から取り除かれる場合に起こり得る。これにより、ICの間にICBUを最大値にセットすることによって避けようとした状況が生じることになる。すなわち、値が最大値ではなかった場合、ページへのその後の更新が、そのICBがセットされなかったため、次の複写動作で捕捉されないことになる。

【0120】この解決法では、ページをS（共用）ラッチする必要がある。これにより、更新トランザクションがページへの更新を完了したことが保証される。すなわち、ラッチによって原子的動作がもたらされる。ページの一貫性のあるコピーをとるために、Sラッチが必要とされることに留意されたい。

【0121】ただ1つのICBU_LSNの値が選択されるので、そのような選択は、コピー・セット中のすべてのページをSラッチした後で行われる。

【0122】ここで図24および図25を参照すると、テープ記憶装置にアーカイブ保存された、完全複写および増分式複写動作後のファイルAが示されている。現コピーを記憶するDASDが障害を起こしたために、ファイルAの現状を回復する必要があるものと仮定する。この場合、図15に示したシステム・カタログに従って、最新の完全コピーとその後の増分式コピーの位置が確立される。それはICRF_LSN=463である。これは、そのときそれぞれテープ・ライブラリ・アドレスT001とT002にアーカイブ保存されたページ番号によって完全コピーおよび増分式コピーをマージする

ことに基づいて、ファイルAを回復することを意味する。

【0123】ここで、図26を参照すると、記憶テープからのファイルAのページ・マージされたコピーが示されている。ページの状態を順方向に変化させるために、ログがLSN463から走査され、LSN490でSMPに、LSN504でDページ1に、530でヘッダ・ページに更新を適用する。これにより、図27に示したような最終ファイル状態になる。

10 【0124】ここで、図28および図29を参照すると、少なくとも1ページへの複数の更新の後のファイルおよび構成の状態が示されている。この場合、トランザクションはDページ4と次いでDページ1を更新する。

【0125】Dページ4については、page_LSN(0)がICBU_LSN(519)よりも小さいので、ICB4はSMP中で1にセットされ、SMPはpage_LSN=570にログされ、Dページ4はB4に更新されpage_LSN=620にログされることになる。

20 【0126】Dページ1については、page_LSN(504)がICBU_LSN(519)よりも小さいので、SMPはアクセスされるが、ICB1がすでに1にセットされているので、SMPへの更新は必要でない。次に、Dページ1が値D1に更新され、page_LSN=630にログされる。

【0127】ここで、ICを取るべき場合、SMPに従ったコピー・セットはDページ1およびDページ4となる。

30 【0128】本発明の上記その他の拡張は、頭記特許請求の範囲に記載されるその趣旨および範囲から逸脱することなしに行うことができる。

【0129】

【発明の効果】本発明の実施により、ログ・ベースの段階式コミットTMSのデータベース中の修正されたページの増分式複写に伴うオーバーヘッドを減少させるための方法および手段を提供することができ、アプリケーションの実行を妨げないログ・ベースの段階式コミットTMS内で修正可能なページのデータベースをアーカイブ保存するための方法および手段を提供することができ、修正可能なページのデータベースへのアクセスを共有する複数ログ・ベース段階式コミット・トランザクション管理システムで動作可能な方法および手段を提供することができる。

【図面の簡単な説明】

【図1】従来技術によるログ・ベースTMSの論理構成を示す図である。

40 【図2】従来技術による、REDOおよびUNDOに関する並列トランザクションと、ログ・ベースの段階式コミット・トランザクション管理システム内での障害の発生との関係を示す図である。

【図 3】 データベースを表す小さなファイルと、分類されたシステム構造およびデータ・ページを格納するシステム・カタログの論理編成を示す図である。

【図 4】 データベースの初期ロード中にログが中断されたものと仮定して、始めにデータをロードし必要なその完全コピーを取った後のファイルを示す図である。

【図 5】 完全バックアップのためのシステム・カタログのエントリを示す図である。

【図 6】 様々なデータ・ページへのトランザクションの修正と様々な SMP および LSN の変更を示す図である。

【図 7】 様々なデータ・ページへのトランザクションの修正と様々な SMP および LSN の変更を示す図である。

【図 8】 様々なデータ・ページへのトランザクションの修正と様々な SMP および LSN の変更を示す図である。

【図 9】 様々なデータ・ページへのトランザクションの修正と様々な SMP および LSN の変更を示す図である。

【図 10】 様々なデータ・ページへのトランザクションの修正と様々な SMP および LSN の変更を示す図である。

【図 11】 図 6 ないし図 10 に示した変更を統合したバージョンを示す図である。

【図 12】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 13】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 14】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 15】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 16】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 17】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 18】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 19】 本発明による増分式複写動作に関連する状態

およびページの変更を示す図である。

【図 20】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 21】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 22】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

【図 23】 本発明による増分式複写動作に関連する状態およびページの変更を示す図である。

10 【図 24】 完全複写動作および増分式複写動作がアーカイブ保存された後のテープ記憶装置上のファイル A のコピーを示す図である。

【図 25】 完全複写動作および増分式複写動作がアーカイブ保存された後のテープ記憶装置上のファイル A のコピーを示す図である。

【図 26】 図 24 および図 25 の FC および IC とログ・レコードとを用いた回復後のファイル A を表す図である。

20 【図 27】 図 24 および図 25 の FC および IC とログ・レコードとを用いた回復後のファイル A を表す図である。

【図 28】 少なくとも 1 つのページへの多重更新の後のファイルおよび構成の状態を示す図である。

【図 29】 少なくとも 1 つのページへの多重更新の後のファイルおよび構成の状態を示す図である。

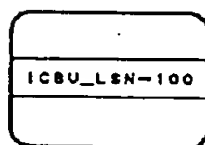
【符号の説明】

- 1 2 トランザクション・プロセス
- 1 3 データベース
- 1 4 データベース管理システム・プログラム
- 1 6 オペレーティング・システム
- 1 8 バッファ・プール
- 2 0 ログ・バッファ
- 2 1 システム・ログ
- 2 2 レコード・マネージャ
- 2 3 バッファ・マネージャ
- 2 4 ログ・マネージャ
- 2 5 回復マネージャ
- 2 6 並列マネージャ
- 3 0 ロック・テーブル

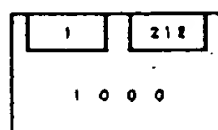
【図 6】

【図 7】

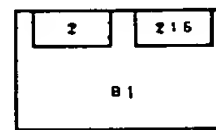
【図 8】



ファイル "A" 用の DBCB

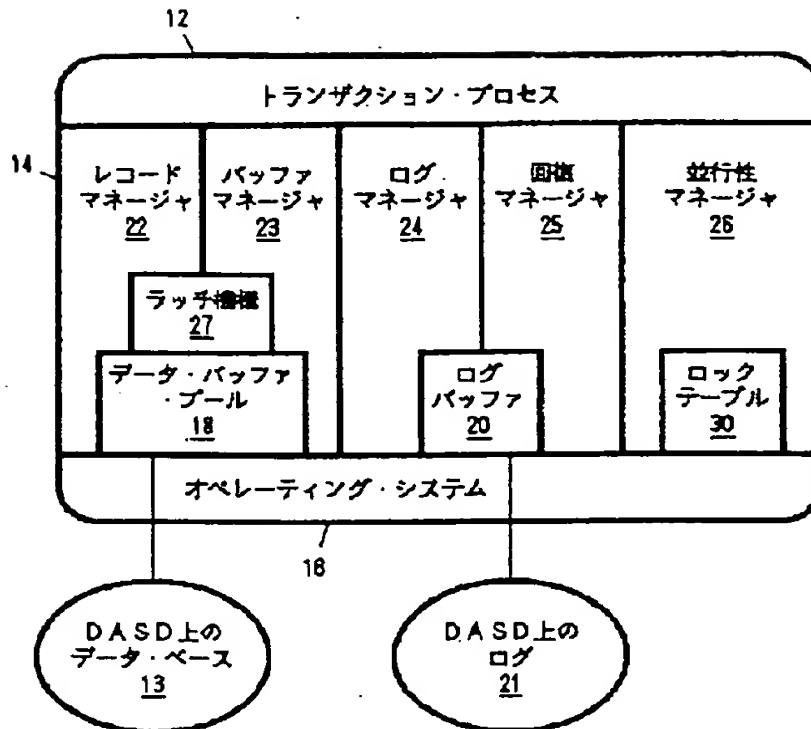


0ページ1に照してセットされた
ICB用のSMPの変更



0ページ1を値A1からB1に
更新する

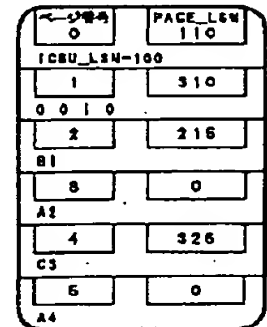
【図1】



【図9】

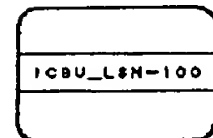


【図11】



Dページ1とDページ3が更新された後のファイルA

【図12】

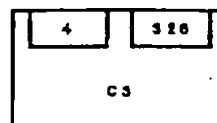
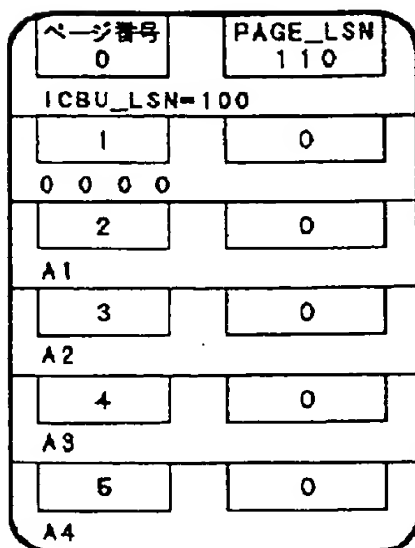


ICB開始時のDBCS

ログ・ベースの段階式コミット・トランザクション
管理システム（従来技術）

【図4】

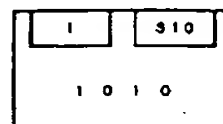
【図10】



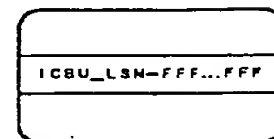
Dページ3をA3からC3に更新する

【図13】

【図14】

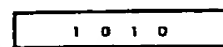


ICB開始時のSMP



最大値にセットされたICBU_LSNを有するDBCS

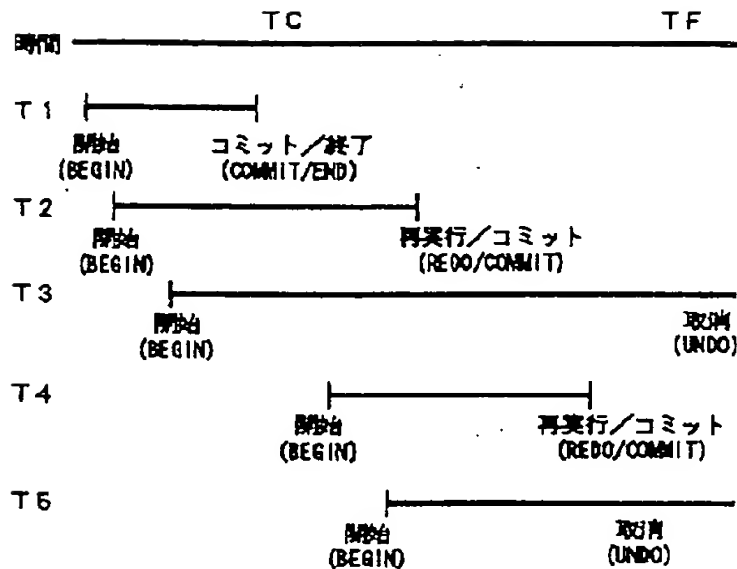
【図17】



ログ複製レコード

初期のロードを行い完全イメージコピーされた後のファイル内のページ

【図2】

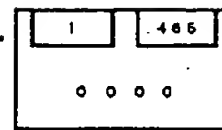


チェックポイント
(時間TC)

システム障害
(時間TF)

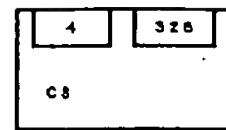
ログ・ベースの段階式コミット・トランザクション管理システム
におけるトランザクション、進行、チェックポイント、障害、
再実行および取消の関係（従来技術）

【図16】



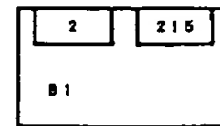
仮記憶域内のSMP

【図21】



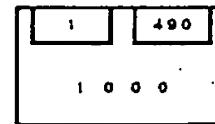
テープT002に書き込まれたDページ3

【図18】



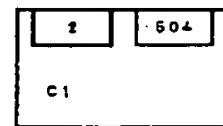
Dページ1

【図19】



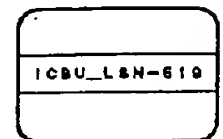
IC後のDページ1更新に必要とするSMP

【図20】



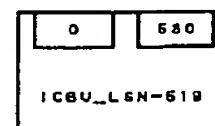
IC復元後に修正されたDページ1

【図22】



ログ頭端にセットされた
ICBU_LSNを有するDBC B

【図23】



新しいICBU_LSNの付与大域を
有する修正されたヘッダ・ページ

【図5】

ファイル名	日付	時間	完全 または 増分式	ファイル・コピーの テープ・アドレス	ICRF_LSN
A	01592	2350	完全	T001	100

完全イメージ・コピー後のシステム・カタログのエントリ

【図3】

ヘッダ・ページ
スペース・マップ・ページ (SMP)
D ページ 1
D ページ 2
D ページ 3
D ページ 4

ファイル "A"

ページ番号 0	PAGE_LSN 0
ICBU_LSN=0	
1	0
ICB1=0 ICB2=0 ICB3=0 ICB4=0	
2	0
A1	
3	0
A2	
4	0
A3	
5	0
A4	

ファイル "A" 内のページ

【図24】

ページ番号 0	PAGE_LSN 000
ICBU_LSN=000 ヘッダ	
1	000
0 0 0 0	SMP
2	000
A1	Dページ1
3	000
A2	Dページ2
4	000
A3	Dページ3
5	000
A4	Dページ4

テープT001上のFIC

ファイル名	日付	時間	完全 または 増分式	ファイル・コピーの テープ・アドレス	ICRF_LSN

システム・カタログ

論理ファイル、ページおよびシステム・カタログの構成

【図25】

ページ番号 0	PAGE_LSN 110
ICBU_LSN=100 ヘッダ	
1	465
0 0 0 0	SMP
2	215
B1	Dページ1
4	326
C3	Dページ2

テープT002上のIIC

【図15】

ファイル名	日付	時間	完全 または 増分式	ファイル・コピーの テープ・アドレス	ICRF_LSN
A	010592	2350	FIC	TD01	100
A	010602	2200	IIC	TD02	483

【図26】

ページ番号 0	PAGE_LSN 110
ICBU_LSN=000 ヘッダ	
1	465
0 0 0 0	SMP
2	215
B1	Dページ1
3	000
A2	Dページ2
4	326
C3	Dページ3
5	000
A4	Dページ4

システム・カタログ

【図27】

ページ番号 0	PAGE_LSN 530
ICBU_LSN=000 ヘッダ	
1	490
1 0 0 0	SMP
2	504
C1	Dページ1
3	000
A2	Dページ2
4	326
C3	Dページ3
5	000
A4	Dページ4

【図28】

ページ番号 0	PAGE_LSN 530
ICBU_LSN=519 ヘッダ	
1	490
1 0 0 0	SMP
2	504
C1	Dページ1
3	000
A2	Dページ2
4	326
C3	Dページ3
5	000
A4	Dページ4

図24と図25のFCと
ICに基づく回復

【図29】

ページ番号 0	PAGE_LSN 530
ICBU_LSN=519 ヘッダ	
1	570
1 0 0 1	SMP
2	630
D1	Dページ1
3	000
A2	Dページ2
4	326
C3	Dページ3
5	620
B4	Dページ4

ICRF_LSN=483 から順方向に
ロールした後の図26のファイル

フロントページの続き

(72)発明者 インデルバル・エス・ナラング
 アメリカ合衆国95070 カリフォルニア州
 サラタガ セラ・オークス・コート13778